Residual convolution LSTM network for machines remaining useful life prediction and uncertainty quantification

Wenting Wang [a], Yaguo Lei [a*], Tao Yan [a], Naipeng Li [a], Asoke K. Nandi [b]

[a] Key Laboratory of Education Ministry for Modern Design & Rotor-Bearing System, Xi'an Jiaotong University, Xi'an 710049, China

[b] Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge UB8 3PH, United Kingdom

*Corresponding author: yaguolei@mail.xjtu.edu.cn

Received Month X, XXXX | Accepted Month X, XXXX | Posted Online Month X, XXXX

Abstract: Recently, deep learning is widely used in the field of remaining useful life (RUL) prediction. Among various deep learning technologies, recurrent neural network (RNN) and its variant, e.g., long short-term memory (LSTM) network, have gained extensive attention for their ability to capture temporal dependence. Although the existing RNN-based methods have demonstrated their RUL prediction effectiveness, they still suffer from the following two limitations: 1) it is difficult for RNN to extract directly degradation features from original monitoring data, and 2) most of the RNN-based prognostics methods are unable to quantify the uncertainty of prediction results. To address the above limitations, this paper proposes a new prognostics method named Residual convolution LSTM (RC-LSTM) network. In RC-LSTM, a new ResNet-based convolution LSTM (Res-ConvLSTM) layer is stacked with convolution LSTM (ConvLSTM) layer to extract degradation representations from monitoring data. Then, predicated on the RUL following a normal distribution, an appropriate output layer is constructed to quantify the uncertainty of the forecast result. Finally, the effectiveness and superiority of RC-LSTM is verified using monitoring data from accelerated degradation tests of rolling element bearings.

Keywords: Deep learning, residual convolution LSTM network, remaining useful life prediction, uncertainty quantification

I. Introduction

Remaining useful life (RUL) prediction, as the effective tool in reducing unplanned shutdowns caused by mechanical failures, is widely utilized in modernized industries to ensure the safety of machines and improve the production efficiency[1][2]. RUL

prediction methods can mainly be divided into two categories[3], i.e., model-based methods and data-driven methods. Model-based methods aim to establish mathematical models to describe the machines degradation process through analyzing the physical failure mechanism. However, in practical cases, failure mechanism-based degradation model is generally difficult to formulate, and the mis-specification of degradation model will seriously impact the accuracy of RUL prediction. Compared with model-based methods, data-driven methods can adaptively model the degradation process without clear physical failure mechanisms through the utilization of machine learning (ML) technologies, such as support vector machine[4] (SVM), gaussian process regression[5] (GPR), artificial neural networks[6] (ANNs), etc. Therefore, data-driven methods have gained increasing attention in recent years.

Nevertheless, the above-mentioned data-driven methods merely use traditional ML methods, which results in limited prediction accuracy. Compared with traditional ML methods, deep learning (DL) methods have more powerful learning capabilities, and possess the capability of establishing more complex mapping relationships[7] between monitoring data and RUL. Therefore, deep learning is now widely used in the field of RUL prediction[8-10]. Among various deep learning technologies, recurrent neural network (RNN)[11][12] and its variant, e.g., long short-term memory (LSTM) network[13-15], are able to effectively capture the time dependence hidden in the degradation process, and have become the promising tool in RUL prediction. Although the existing approaches have demonstrated their RUL prediction effectiveness, they still suffering the following two limitations.

1) Since RNN lacks consideration of spatial correlation[16], it is difficult for RNN to extract directly degradation features from original monitoring data. Therefore, most of the RNN-based prognostic approaches are combined with hand-crafted features, which affect the accuracy and generalization of the forecast results, resulting in the fact that they are accurate only in some specific scenarios.
2) The uncertainty quantification of prognostic result is pivotal in making maintenance decisions, but most of the RNN-based RUL prediction methods only provide a point estimation. In the case of complex working conditions, the stability of RUL prediction results is often compromised, which would reduce the credibility of the method in providing guidance for predictive maintenance scheduling. Furthermore, the existing DL-based uncertain quantification methods mainly include two types: bootstrap[17] and monte-carlo-dropout[18] (MC-dropout). Recently, Jason et al.[19] and Liu et al.[20] quantified the uncertainty of predicted RUL with bootstrap method. However, bootstrap-based methods require resampling the original training data and is time-consuming. Peng et al.[21] and Wang et al.[22] combined MC-dropout with different neural networks to estimate the uncertainty. But MC-dropout-based methods need to run multiple times during the test stage, which is not efficient enough and requires additional computational burden.

To address the above-mentioned limitations, this paper proposes a new prognostic method named residual convolution long short-term memory (RC-LSTM) network for the RUL predictions of machines. In the proposed method, a ResNet-based convolution long short-term memory (Res-

ConvLSTM) layer, which is improved from convolution long short-term memory (ConvLSTM) network[16], is first built to extract directly degradation representations and capture time dependence information from monitoring data. Then, an ordinary ConvLSTM layer is used to extract further degradation information. After that, predicated on the RUL following a normal distribution, an appropriate output layer is constructed to directly quantify the uncertainty of the forecast result without additional computational burden. Finally, the superiority of the proposed Res-ConvLSTM is validated using vibration data from accelerated degradation tests of rolling element bearings.

II. The proposed method

The framework of RC-LSTM network is shown in Fig.1. First, Res-ConvLSTM layer and ConvLSTM layer are stacked with Batch normalization (BN) layer[23] to extract informative degradation representations. Then, the representations are input into normal distribution output layer to obtain the predicted RUL and to quantify uncertainty.

In this section, we first introduce the ConvLSTM unit. Then, Res-ConvLSTM, where a new core building unit, improved from ConvLSTM, is further detailed. Finally, the normal distribution output layer and its loss function are described in details.

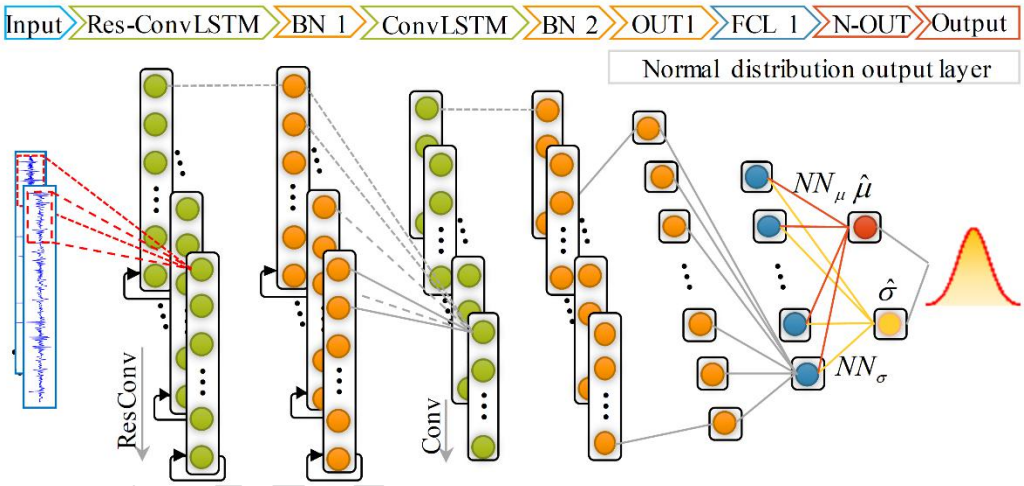

Fig. 1. Architecture of the proposed RC-LSTM.

A. ConvLSTM unit

It has been justified that LSTM lacks consideration of spatial correlation, while in the meantime, CNN owns the capability of extracting spatial features. Accordingly, it is considered to embed CNN into LSTM to construct ConvLSTM unit to capture time series information and extract degradation representations simultaneously. The structure of ConvLSTM is shown in Fig. 2.

Given the current input $\boldsymbol{x}_t$, the hidden state $\boldsymbol{h}_{t-1}$ and the cell state $\boldsymbol{c}_{t-1}$ at previous moment, the output representation $\boldsymbol{h}_t$ can be calculated as follows

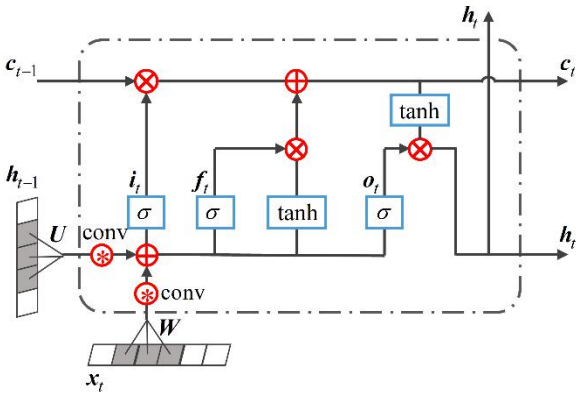

Fig. 2. The structure of ConvLSTM unit.

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}^{(i)} * \boldsymbol{x}_t + \boldsymbol{U}^{(i)} * \boldsymbol{h}_{t-1}) \quad (1)$$

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}^{(f)} * \boldsymbol{x}_t + \boldsymbol{U}^{(f)} * \boldsymbol{h}_{t-1}) \quad (2)$$

$$\boldsymbol{c}'_t = \tanh(\boldsymbol{W}^{(c)} * \boldsymbol{x}_t + \boldsymbol{U}^{(c)} * \boldsymbol{h}_{t-1}) \quad (3)$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \cdot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \cdot \boldsymbol{c}'_t \quad (4)$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}^{(o)} * \boldsymbol{x}_t + \boldsymbol{U}^{(o)} * \boldsymbol{h}_{t-1}) \quad (5)$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \cdot \tanh(\boldsymbol{c}_t) \quad (6)$$

where $\boldsymbol{i}_t$, $\boldsymbol{f}_t$, $\boldsymbol{o}_t$ are input gate, forget gate and output gate respectively. $*$ represents the convolution operator, $\boldsymbol{c}'_t$ is the current input, $\boldsymbol{c}_t$ is the current cell state, and $\boldsymbol{W}$, $\boldsymbol{U}$ are convolution kernels of different gates.

Compared with the traditionally used LSTM, the main difference of ConvLSTM is that the fully connected structure in LSTM is replaced with the convolutional structure, which comprehensively utilizes the advantages of LSTM and CNN to extract simultaneously the local degradation feature and time dependence.

B. Res-ConvLSTM unit

As shown in Fig. 2, ConvLSTM uses single layer convolution to extract local features from the current input $\boldsymbol{x}_t$. However, for complex original monitoring data, single layer convolution is not sufficient to extract sensitive degradation representations. To address the above-mentioned problem, this paper constructs a Res-ConvLSTM unit, which utilizes deep residual convolutional neural network (Resnet) to replace single layer convolution. The structure of Res-ConvLSTM unit is shown in Fig. 3.

Compared to ordinary convolutional neural networks, Resnet utilizes an identity mapping to address the problem of vanishing gradient and gradient explosion, which ensures the stability of network training when increasing the depth of network. Moreover, the structure of basic Resnet unit is shown in Fig. 4.

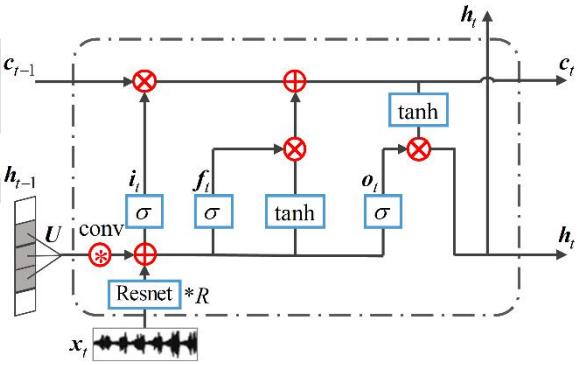

Fig. 3. The structure of Res-ConvLSTM unit.

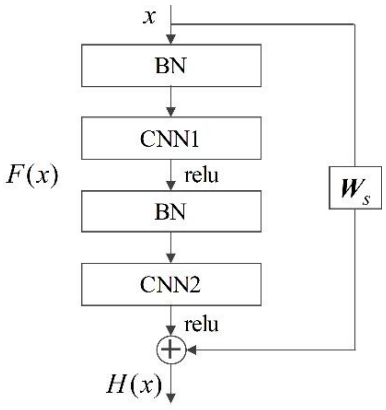

Fig. 4. The structure of basic Resnet unit.

Given the input vector $\boldsymbol{x}$, The output $H(\boldsymbol{x})$ of Resnet unit is calculated as follow

$$H(\boldsymbol{x}) = F(\boldsymbol{x}) + \boldsymbol{W}_s \boldsymbol{x} \tag{7}$$

Where $F(\boldsymbol{x})$ is the output that ignores the identity mapping, and $\boldsymbol{W}_s$ is a linear projection weight matrix, which aims to maintain the uniform dimensions of $\boldsymbol{x}$ and $F(\boldsymbol{x})$.

Hence, by stacking R Resnet units to substitute the single layer convolution, Res-ConvLSTM unit has greatly improved its local feature extraction capabilities, thereby realizing effective processing of massive monitoring data.

C. Normal distribution output layer

Different from the traditional point estimation-based methods, this paper assumes that the predicted RUL follows a normal distribution, and correspondingly constructs the normal distribution output layer to quantify uncertainty. The constructed layer is shown in Fig. 1, which exploits two different fully connected layers (FCLs) NN_μ and NN_σ to obtain mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$. Among them, the mean $\hat{\mu}$ denotes the mean value of predicted RUL, and the standard deviation $\hat{\sigma}$ represents the uncertainty of the predicted result. Based on the above results, the conditional probability of RUL is obtained as

$$P(y \mid \boldsymbol{x}, NN_\mu, NN_\sigma) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} e^{-\frac{(y-\hat{\mu})^2}{2\hat{\sigma}^2}} \tag{8}$$

where y is the actual RUL, $\boldsymbol{x}$ is the input of normal distribution output layer. Consequently, according to the mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$, the prediction confidence interval (PI) corresponding to $(1-\alpha)\%$ can be obtained as follow

$$\mathrm{PI} = \hat{\mu} \pm z_{1-\alpha/2}\hat{\sigma} \tag{9}$$

Furthermore, since the standard deviation $\hat{\sigma}$ is an unknown prior, the supervised learning method is incapable of adjusting the parameters in NN_σ. In view of this, the maximum likelihood estimation method is utilized to optimize the network parameters. Correspondingly, the log-likelihood loss function (LLF) L_1 is formulated as

$$L_1 = \frac{1}{2}\sum_{i=1}^{n}[\ln(\hat{\sigma}_i^{\,2}) + \frac{(y_i - \hat{\mu}_i)^2}{\hat{\sigma}_i^{\,2}}] \tag{10}$$

where n is batch size. During the training process, the goal is to find the optimal $\hat{\mu}$ and $\hat{\sigma}$ that corresponds to the ascent direction of actual RUL condition probability. However, it is difficult to optimize simultaneously both mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ for L_1. Moreover, in some cases when actual RUL is known, another feasible option is to utilize mean square error (MSE) to optimize $\hat{\mu}$ to ensure the accuracy of the predicted mean, then LLF can be further optimized with respect to both $\hat{\mu}$ and $\hat{\sigma}$. The L_2 loss function of the above described optimization process is defined as

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2 \tag{11}$$

$$L_2 = k_2^1 \mathrm{MSE} + k_2^2 \mathrm{LLF} \tag{12}$$

where k_2^1 and k_2^2 are the weights corresponding to MSE and LLF, which are adjusted with the training epoch.

In addition, it is a commonsense that the more light-tailed the distribution of PI is, the more effective the maintenance scheduling can be. Based on this premise, the prediction interval averaged width[24] (PIAW) indicator is introduced into L_2 loss function to further clarify the optimization direction of standard deviation $\hat{\sigma}$, The L_3 loss function is correspondingly defined as follows

$$\text{PIAW}=\frac{1}{n}\sum_{i=1}^{n}(U_i - L_i) \tag{13}$$

$$L_3 = k_3^1\text{MSE} + k_3^2\text{LLF} + k_3^3\text{PIAW} \tag{14}$$

where U_i and L_i are the upper and lower limits of PI, respectively, k_3^1 , k_3^2 and k_3^3 are the weights corresponding to MSE, LLF and PIAW.

III. Case study: RUL prediction of rolling element bearings

To verify the effectiveness of the proposed RC-LSTM, vibration data collected from accelerated degradation tests of rolling element bearings are used, and four state-of-the-art prognostic approaches are compared in this section.

A. Data description

The datasets used in this paper are the XJTU-SY Bearing Datasets[3]. And the accelerated degradation tests of rolling element bearings are conducted in the testbed as shown in Fig. 5, which consists of an alternating current (AC) motor, a rotating shaft, a support bearing, a test bearing, a hydraulic loading system, a motor speed controller, etc. By controlling the speed and load, the testbed is capable of conducting degradation tests of bearings under different operating conditions.

As recorded in Table 1, fifteen LDK UER204 bearings were tested under three different operating conditions. During each test, two PCB 352C33 acceleration sensors were mounted in the horizontal and vertical directions of the test bearing to monitor the degradation process of the rolling bearing. And during the experiments, the sampling frequency was set to be 25.6 kHz, with the sampling time of 1.28 s and the sampling interval of 60 s. Put otherwise, 32768 data points can be obtained per sampling. In this section, Bearing1_4, Bearing2_4, Bearing3_4 and Bearing1_5, Bearing2_5, Bearing3_5 are selected as the testing dataset, respectively. And the remaining bearings are assigned as the training dataset.

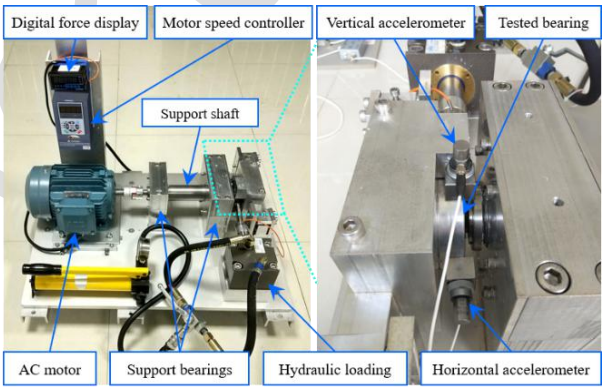

Fig. 5. Testbed of rolling element bearings.

Table 1. Dataset description

Radial force	Rotating speed/rpm	Bearing datasets
12 kN	2100 rpm	Bearing1_1 Bearing1_2 Bearing1_3 **Bearing1_4 Bearing1_5**
11 kN	2250 rpm	Bearing2_1 Bearing2_2 Bearing2_3 **Bearing2_4 Bearing2_5**
10 kN	2400 rpm	Bearing3_1 Bearing3_2 Bearing3_3 **Bearing3_4 Bearing3_5**

B. Prognostic metrics

In this part, apart from the two commonly used prognostic metrics, RMSE and score function[25], two other metrics, i.e., $\alpha-\lambda$

accuracy[26] ($A_{\alpha-\lambda}$) and average interval score[27] (AIS), are also utilized to evaluate quantitatively the prediction performance. $A_{\alpha-\lambda}$ is a binary metric that evaluates whether the prediction result falls within $\alpha-$bounds at time λ. In this paper, α and λ are set to 0.3 and 0.5. The AIS is employed to evaluate the comprehensive performance of the prediction interval, which is defined as follows

$$\xi_i^{(\alpha)} = U_i^{(\alpha)} - L_i^{(\alpha)} \tag{15}$$

$$S_i^{(\alpha)} = \begin{cases} -2\alpha\xi_i^{(\alpha)} - 4(L_i^{(\alpha)} - y_i), & if\ y_i < L_i^{(\alpha)} \\ -2\alpha\xi_i^{(\alpha)}, & if\ L_i^{(\alpha)} < y_i < U_i^{(\alpha)} \\ -2\alpha\xi_i^{(\alpha)} - 4(y_i - U_i^{(\alpha)}), & if\ y_i > U_i^{(\alpha)} \end{cases} \tag{16}$$

$$\text{AIS} = \frac{1}{n}\sum_{i=1}^{n} S^{(\alpha)} \tag{17}$$

where α is the confidence level, ξ is the interval length of the *i*-th PI, and S is the corresponding interval score that will impose punishment if the actual RUL is outside the PI. According to this definition, AIS is always a negative value that shares the positive correlation along with change of the convergence rate.

C. Configuration of RC-LSTM

Firstly, time window embedding is conducted on the monitoring data. Then, in RC-LSTM, the hyperparameters determined by grid search is summarized in Table 2, and Adam optimizer is adopted to minimize the loss function. In addition, the weights in different loss functions are shown in Table 3.

Table 2. Configuration of RC-LSTM

Hyperparameter	Size	Hyperparameter	Size
Number of Resnet units	9	Number of neurons in NN_μ and NN_σ	100
Number of kernels in Res-ConvLSTM	256	Number of kernels in ConvLSTM	256
Kernels size in Res-ConvLSTM	3×1	Kernels size in ConvLSTM	3×1
Time step	7	epoch	100
Batch size	64	Learning rate	0.001

Table 3. Weights in different loss functions

Epoch	L_2	L_3
[0, 30]	$k_2^1=0.8, k_2^2=0.2$	$k_3^1=0.8, k_3^2=0.2, k_3^3=0$
(30, 60]	$k_2^1=0.5, k_2^2=0.5$	$k_3^1=0.5, k_3^2=0.5, k_3^3=0.1$
(60, 80]	$k_2^1=0.3, k_2^2=0.7$	$k_3^1=0.3, k_3^2=0.7, k_3^3=0.6$
(80, 100]	$k_2^1=0.1, k_2^2=0.9$	$k_3^1=0.1, k_3^2=0.9, k_3^3=1$

D. RUL prediction for Bearings

In this section, the advantages from the proposed loss function are first investigated and discussed. Then, the proposed RC-LSTM is compared with other state-of-the-art prognostics methods to demonstrate its superiority.

1. Discussion of loss function

In addition to the ordinary log-likelihood loss function L_1 defined in , this paper proposes two different loss functions, i.e., L_2 defined in and L_3 defined in . To illustrate the advantages of L_2 and L_3, all three loss functions are used to predict the RUL of bearings. Fig. 6 shows the prediction results of Bearing 3_5 under different loss functions. Meanwhile, the performance estimation results under all three scenarios are tabulated in Table 4.

Table 4. The performance estimation results of these three loss functions

Metrics		L_1	L_2	L_3
Bearing 1_5	RMSE	8.04	5.91	**5.78**
	Score	22.28	18.73	**17.64**
	AIS	-23.21	-21.83	**-19.96**
	$A_{\alpha-\lambda}$	53.48%	69.44%	**71.57%**
Bearing 2_5	RMSE	30.21	29.40	**21.77**
	Score	3358.18	3165.99	**1348.43**

	AIS	-131.65	-123.88	**-115.85**
	$A_{\alpha-\lambda}$	52.69%	53.30%	**64.67%**
	RMSE	5.85	5.35	**4.31**
Bearing	Score	32.53	25.24	**20.56**
3_5	AIS	-35.97	-33.42	**-27.69**
	$A_{\alpha-\lambda}$	59.26%	64.81%	**68.52%**

It can be observed from Table 4 and Fig. 6 that in the RUL prediction of bearings, compared with ordinary log-likelihood loss function L_1, both L_2 and L_3 achieve lower RMSE, Score values and higher $A_{\alpha-\lambda}$ values, which indicates that by introducing MSE and prioritizing the optimization of prediction mean $\hat{\mu}$, the performance of network can be effectively improved, and the mean value $\hat{\mu}$ of RUL prediction results can be obtained with higher accuracy. Moreover, it can also be clearly seen that by considering the directional optimization of prediction standard deviation $\hat{\sigma}$, the PI corresponding to L_3 is significantly narrower than that of L_1 and L_2, which means L_3 is able to reduce sufficiently the uncertainty of prediction results, and provide a more applicable PI. As a result, benefiting from introducing the MSE and PIAW, the proposed loss function effectively improves the performance of prognostics model.

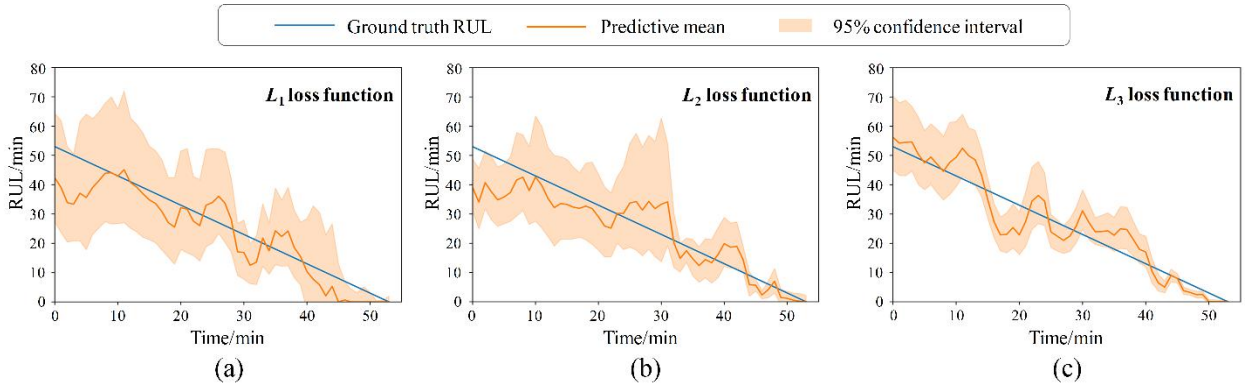

Fig. 6. RUL prediction results of Bearing3_5 under different loss functions. (a) L_1 loss function. (b) L_2 loss function. (c) L_3 loss function.

2. Comparison with the state-of-the-art prognostics methods

In this section, three other state-of-the-art point estimation prognostics approaches are firstly implemented to predict RUL for bearings for comparison, which are named as N1, N2 and N3, respectively. Among them, N1[28] first converts the time-domain signal into the frequency-domain, then inputs it into one-dimensional CNN and simple recurrent unit network to realize RUL prediction. N2[29] utilizes LSTM and attention mechanism to learn degradation representations from original monitoring data and predict RUL. N3[30] is a prognostics method that combines CNN, LSTM and attention mechanism. Table 5 summarizes the evaluation metrics of different networks.

The result shows that the proposed RC-LSTM gets the lowest RMSE values in all cases and the lowest Score values in all cases but one. This signifies the superiority of RC- LSTM, compared to the three state of the art methods in RUL prediction accuracy. More importantly, RC-LSTM is capable of providing a probabilistic distribution to quantify uncertainty, which overcomes the

Table 5. The performance estimation results of different networks

Bearing dataset	N1		N2		N3		Proposed	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
Bearing1_4	9.76	100.84	9.35	89.79	9.21	82.91	**7.73**	**58.61**
Bearing2_4	14.67	218.11	17.89	363.04	15.41	285.24	**12.86**	**184.15**
Bearing3_4	36.66	6151.82	44.75	7576.17	37.57	6503.24	**31.23**	**4846.75**
Bearing1_5	5.98	18.70	6.52	20.77	6.04	**14.34**	**5.78**	17.64
Bearing2_5	31.31	4036.24	32.89	2158.24	27.39	2467.59	**21.77**	**1348.43**
Bearing3_5	5.85	32.10	6.48	37.56	5.23	25.17	**4.31**	**20.56**

limitation of traditional point estimation prognostics approaches. Therefore, the prediction performance of RC-LSTM is more useful in applications than the other three methods.

To further verify the performance of the proposed method, we also compare the proposed method with other DL-based uncertainty quantification methods, including Bayesian multiscale CNN-based method[21] (BMSCNN) and Bayesian recurrent convolutional neural network[22] (BRCNN). Among them, BMSCNN combine Monte-Carlo-dropout with deep multiscale CNN to achieve uncertainty quantification. BRCNN first constructs a network structure named recurrent convolutional neural network, and then utilizes MC-dropout to quantify the uncertainty. The comparison results are tabulated in Table 6.

Table 6. The performance estimation results of different DL-based uncertainty quantification methods

Bearing dataset	BMSCNN				BRCNN				Proposed			
	RMSE	Score	$A_{\alpha-\lambda}$ / %	AIS	RMSE	Score	$A_{\alpha-\lambda}$ / %	AIS	RMSE	Score	$A_{\alpha-\lambda}$ / %	AIS
Bearing1_4	11.66	84.39	43.33	-153.39	9.68	92.24	56.67	-83.90	**7.73**	**58.61**	**66.67**	**-49.66**
Bearing2_4	18.81	450.61	42.50	-191.00	13.41	**168.89**	55.00	-87.54	**12.86**	184.15	**61.25**	**-69.36**
Bearing3_4	46.41	7437.45	31.85	-345.36	40.60	6518.99	39.49	-273.61	**31.23**	**4846.75**	**45.22**	**-224.66**
Bearing1_5	9.15	30.76	47.83	-31.38	6.28	21.82	66.09	-29.99	**5.78**	**17.64**	**71.57**	**-19.96**
Bearing2_5	27.30	4767.90	46.71	-141.27	24.06	2204.14	58.68	-127.24	**21.77**	**1348.43**	**64.67**	**-115.85**
Bearing3_5	6.59	35.68	55.56	-37.00	5.54	26.98	64.81	-34.01	**4.31**	**20.56**	**68.52**	**-27.69**

From this table, it can be clearly seen that the proposed method gets lower score, RMSE values and higher $A_{\alpha-\lambda}$ values, which indicates the proposed RC-LSTM achieves higher prediction accuracy than BMSCNN and BRCNN. Furthermore, the AIS value of the proposed method is significantly higher than other uncertainty quantification methods, this means that the proposed RC-LSTM achieves a better trade-off between the prediction interval coverage and the interval width. Based on the above analyses, the RC-LSTM has a slightly better performance than other DL-based uncertainty quantification methods in RUL prediction of bearings.

IV. Conclusion

This paper has proposed a new prognostics method named RC-LSTM to predict the RUL of machines. In RC-LSTM, a residual convolution long short-term memory layer is constructed to extract degradation

representations from monitoring data. Then, a convolution long short-term memory layer is stacked to capture further time dependence. After that, through constructing normal distribution output layer and improving loss function, the proposed method is able to quantify effectively the uncertainty of forecast results. Finally, the vibration data of bearings has been employed to evaluate the proposed RC-LSTM, and the forecast results have been compared with other state-of-the-art prognostics methods. Experimental results indicate the effectiveness and superiority of the proposed method. Moreover, different from the traditional point estimation prognostics approaches, RC-LSTM can provide probabilistic prediction results, which facilitates making effective maintenance decisions.

Acknowledgment

This research was supported by National Natural Science Foundation of China (52005387, 52025056), Project funded by China Postdoctoral Science Foundation(2020M673380), and Fundamental Research Funds for the Central Universities.

References

[1] Lei Y, Li N, Guo L, et al. Machinery health prognostics: A systematic review from data acquisition to RUL prediction[J]. Mechanical systems and signal processing, 2018, 104: 799-834.

[2] Dong H, Hou N, Wang Z, et al. Finite-horizon fault estimation under imperfect measurements and stochastic communication protocol: Dealing with finite-time boundedness[J]. International Journal of Robust and Nonlinear Control, 2019, 29(1): 117-134.

[3] Wang B, Lei Y, Li N, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings[J]. IEEE Transactions on Reliability, 2018, 69(1): 401-412.

[4] Khelif R, Chebel-Morello B, Malinowski S, et al. Direct remaining useful life estimation based on support vector regression[J]. IEEE Transactions on industrial electronics, 2016, 64(3): 2276-2285.

[5] Kong D, Chen Y, Li N. Gaussian process regression for tool wear prediction[J]. Mechanical systems and signal processing, 2018, 104: 556-574.

[6] Lim P, Goh C K, Tan K C. A novel time series-histogram of features (TS-HoF) method for prognostic applications[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018, 2(3): 204-213.

[7] Yang F, Dong H, Wang Z, et al. A new approach to non-fragile state estimation for continuous neural networks with time-delays[J]. Neurocomputing, 2016, 197: 205-211.

[8] Wang B, Lei Y, Li N, et al. Multiscale Convolutional Attention Network for Predicting Remaining Useful Life of Machinery[J]. IEEE Transactions on Industrial Electronics, 2020, 68(8): 7496-7504.

[9] Hu C H, Pei H, Si X S, et al. A prognostic model based on DBN and diffusion process for degrading bearing[J]. IEEE Transactions on Industrial Electronics, 2019, 67(10): 8767-8777.

[10] Kara A. A data-driven approach based on deep neural networks for lithium-ion battery prognostics[J]. Neural Computing and Applications, 2021: 1-14.

[11] Yu W, Kim I I Y, Mechefske C. An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme[J]. Reliability Engineering & System Safety, 2020, 199: 106926.

[12] Hua Z, Zheng Z, Pahon E, et al. Remaining useful life prediction of PEMFC systems under dynamic operating conditions[J]. Energy Conversion and Management, 2021, 231: 113825.

[13] Zhou J T, Zhao X, Gao J. Tool remaining useful life prediction method based on LSTM under variable working conditions[J]. The International Journal of Advanced Manufacturing Technology, 2019, 104(9): 4715-4726.

[14] Wu J, Hu K, Cheng Y, et al. Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network[J]. ISA transactions, 2020, 97: 241-250.

[15] Wu J Y, Wu M, Chen Z, et al. Degradation-aware remaining useful life prediction with LSTM autoencoder[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-10.

[16] Xingjian S H I, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[C]//Advances in neural information processing systems. 2015: 802-810.

[17] Efron B, Tibshirani R J. An introduction to the bootstrap[M]. CRC press, 1994.

[18] Gal Y, Ghahramani Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning[C]//international conference on machine learning. PMLR, 2016: 1050-1059.

[19] Deutsch J, He D. Using deep learning-based approach to predict remaining useful life of rotating components[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017, 48(1): 11-20.

[20] Liu D, Xie W, Liao H, et al. An integrated probabilistic approach to lithium-ion battery remaining useful life estimation[J]. IEEE Transactions on Instrumentation and Measurement, 2014, 64(3): 660-670.

[21] Peng W, Ye Z S, Chen N. Bayesian deep-learning-based health prognostics toward prognostics uncertainty[J]. IEEE Transactions on Industrial Electronics, 2019, 67(3): 2283-2293.

[22] Wang B, Lei Y, Yan T, et al. Recurrent convolutional neural network: A new framework for remaining useful life prediction of machinery[J]. Neurocomputing, 2020, 379: 117-129.

[23] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. PMLR, 2015: 448-456.

[24] Khosravi A, Nahavandi S, Creighton D. A prediction interval-based approach to determine optimal structures of neural network metamodels[J]. Expert systems with applications, 2010, 37(3): 2377-2387.

[25] Saxena A, Goebel K, Simon D, et al. Damage propagation modeling for aircraft engine run-to-failure simulation[C]//2008 international conference on prognostics and health management. IEEE, 2008: 1-9.

[26] Saxena A, Celaya J, Saha B, et al. Metrics for offline evaluation of prognostic performance[J]. International Journal of Prognostics and health management, 2010, 1(1): 4-23.

[27] Ferguson T S. Mathematical statistics: A decision theoretic approach[M]. Academic press, 2014.

[28] Yao D, Li B, Liu H, et al. Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit[J]. Measurement, 2021, 175: 109166.

[29] Chen Z, Wu M, Zhao R, et al. Machine remaining useful life prediction via an attention-based deep learning approach[J]. IEEE Transactions on Industrial Electronics, 2020, 68(3): 2521-2531.

[30] Jiang J R, Lee J E, Zeng Y M. Time series multiple channel convolutional neural network with attention-based long short-term memory for predicting bearing remaining useful life[J]. Sensors, 2020, 20(1): 166.